**Wang, D; Murphy, M (2002) Trends and differentials in menarcheal age in China. Journal of biosocial science, 34 (3). pp. 349-61. ISSN 0021-9320 DOI: https://doi.org/10.1017/S0021932002003498**



DOI: 10.1017/S0021932002003498


# TRENDS AND DIFFERENTIALS IN MENARCHEAL AGE IN CHINA


DUOLAO WANG and MICHAEL MURPHY



Journal of Biosocial Science / Volume 34 / Issue 03 / July 2002, pp 349 - 361
DOI: DOI:10.1017/S0021932002003498, Published online: 26 July 2002

*J. biosoc. Sci.* (2002) **34**, 349–361


# TRENDS AND DIFFERENTIALS IN MENARCHEAL AGE IN CHINA

DUOLAO WANG* AND MICHAEL MURPHY†

**Medical Statistics Unit, Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT and †Population Studies, Department of Social Policy, London School of Economics, Houghton Street, London WC2A 2AE*

**Summary**. This study examines trends in menarcheal age of women born in China between 1950 and 1973, and explores the impact of relevant social background characteristics on the timing of first menarche. Data on recalled ages of menarche collected in the 1988 Chinese Two-per-Thousand Fertility Survey are used in a linear regression model where the covariates are transformed with the help of an Alternating Conditional Expectation (ACE) algorithm. The results indicate that a trend towards early menarche has evolved in China during recent decades. The pattern of early menarche is especially pronounced among women residing in urban areas, and those who are better educated.

## Introduction

The age at which a girl attains menarche – the appearance of first menstruation – defines a particular stage of her maturation. Age at menarche is an indicator of female physiological development, health, timing of maturation and nutritional status, and a predictor of fecundity and risk for miscarriage and unsuccessful pregnancy outcomes (Tanner & Eveleth, 1975; Urdy & Cliquet, 1982; Wyshak, 1983; Sandler, Wilcox & Honey, 1984; Frisch, 1985; Eveleth, 1986; Hediger & Stine, 1987). There is strong evidence of a downward secular trend in Western countries during the nineteenth century and the decline stabilized in the twentieth century (Wyshak & Frisch, 1982; Eveleth, 1986; Tanner & Eveleth, 1975). However, in China, the mean age at menarche has only started to approach that of Western countries in recent decades (Graham, Larsen & Xu, 1999; Lin *et al.*, 1992). As China is a huge country with consequent large variation in living conditions, economic developments, and general health status in different regions, more studies are required to assess the determinants of age at menarche.

There have been some studies on menarcheal age of Chinese females, most of which have been restricted to local surveys (Quan, 1984; Lin & Yuan, 1989; Graham *et al.*, 1999). However, estimates of trends in the average age at menarche for the whole nation, as well as for particular subgroups of the population, are very rare. Little is also known about socioeconomic differentials in the timing of menarcheal age in China. An exception is the study of Lin *et al.* (1992) of differentials in menarcheal age in China using a sample of 200,000 Chinese girls aged 7–22 years from a nationally representative survey. However, this study examined mainly regional and ethnic differences in age of menarche but it could not explore time trends because information was collected only from young women. Taken as a whole, these studies show a generally declining age at menarche in China. However, they typically cover different geographical units, the population used may not be randomly selected and they use different collection methods. In this study, national-level trends are derived from a single source, which should ensure greater consistency in reporting. The levels found are within the ranges found in other studies.

This paper examines trends and differentials of menarcheal age of cohorts born in China between 1950 and 1973 using a large nationally representative survey. The ACE technique for transformation of variables is used in the linear regression analysis of recalled age at menarche. This approach overcomes the general problem in multiple regression analysis of finding the appropriate way of simultaneously coding the response and independent variables so that the relationship between them is best described, rather than coding such variables in an arbitrary way.

## Materials and methods

The data for this study come from a 10% subset of China's State Family Planning Commission in 1988 'Two-per-Thousand Fertility Survey' of 0·2% of ever-married women aged 15–57 resident in mainland China, which used a single-stage non-proportionate stratified-cluster sampling design (Lavely, 1991). The central elements of the survey were pregnancy and contraceptive use histories, but data on recalled age at menarche were also collected. The sample for this analysis consists of 27,655 ever-married women aged between 15 and 38, that is those born between 1950 and 1973, making it possible to examine trends in menarcheal age in recent decades. The sample is of ever-married women and therefore the youngest cohorts become progressively selected for young age at marriage, since, for example, women in the survey aged 20 must have been married by this age whereas older women could have been married at older ages. Girls who marry early are likely to have had earlier average age at menarche since not only is it a requirement for marriage, but also it is associated with earlier sexual maturity. Such a correlation would tend to reinforce any observed decline in age at menarche. However, there is actually a small negative correlation, $r = -0{\cdot}04$, between age at menarche and age at marriage for women born before 1965 whose average age at marriage was over 21·5 years, and only a small positive correlation, $r = 0{\cdot}10$, for women born later whose average age at marriage was 19·6 years. The negative correlation is possibly due to selection effects, in that groups that tend to marry young, such as rural women, have older ages at menarche. The weak relationship observed between age of menarche and marriage is consistent with that of Udry & Cliquet (1982), who found less association between them for Malaysian Chinese than for any of the other groups they studied. Thus, the selectivity issue does not appear to alter the interpretation of the results. To further assess the sensitivity of the findings to age at marriage, the regression equations discussed later were re-estimated, but confined to women married above age 20, and it was found that the results remained essentially unaltered.

A common problem in using data from retrospective studies is that they may suffer from recall errors on the dating of events, particularly for events that took place further back in time. However, an unusually high accuracy in reporting age-related events in recent Chinese fertility surveys and censuses has been documented (Coale & Banister, 1994; Wang & Murphy, 1998).

To estimate the effects of socioeconomic variables on recalled age at menarche, a multiple linear regression model is used. Selection of relevant independent variables is guided by previous studies on the determinants of menarcheal age and constrained by those that were included in the survey. Therefore the following variables were selected: year of birth, urban or rural residence, education, occupation, geographic region of residence, and ethnicity. Table 1 shows descriptive statistics for the variables included in this analysis. Women's education and occupation were measured at the time of survey. Although ideally parental education and occupation should be used to investigate the socioeconomic impact of such variables on the age at menarche, it is still expected that the two variables included will capture some socioeconomic variability in age at menarche.

Variables can be classified into two main types: (1) categorical (ordered or unordered) with multiple categories such as geographic region of residence or educational attainment, and those that are binary variables such as residence (urban/rural) or ethnicity (Han majority/minority); (2) numeric variables – discrete, interval or continuous – such as the woman's year of birth. For binary variables, a simple indicator or dummy variable can be used to represent its effect in a regression model. For coding non-binomial categorical and numeric variables, the ACE algorithm described in the next section is used.

### *ACE-guided transformation of covariates*

In linear regression analysis, the objective of fully exploring and explaining the effect of covariates on the dependent variable of age at menarche is facilitated by properly coding the independent variables. For both categorical and numerical variables, a common technique is to create a set of dummy variables, but this is often done on an *ad hoc* basis. It may be possible to collapse some neighbouring groups, but it is not known how to find the most efficient cutting points. A numeric variable may be entered into a linear regression model as a simple linear form or as a transformed one, such as a polynomial form with linear and squared terms, or a non-linear function that cannot be estimated directly. However, there is usually no prior knowledge or consensus on the appropriate form to be used.

Here, efficient transformations are searched for using the ACE (Alternating Conditional Expectation) algorithm (Breiman & Friedman, 1985; De Veaux, 1989; Wang & Murphy, 1998), a technique for finding non-parametric transformations of both the independent and dependent variables in multiple linear regression such that the relationship between them is as linear as possible. The ACE results can be used

**Table 1.** List of covariates with descriptive statistics

| Covariate | Statistics |
|---|---|
| Year of birth (1950–1973, mean) | 1958·7 |
| Place of residence (%) | |
| Rural | 76·5 |
| Urban | 23·5 |
| Educational attainment (%) | |
| Illiterate/semiliterate | 30·7 |
| Primary school | 26·8 |
| Junior middle | 27·8 |
| Senior middle | 13·5 |
| College and above | 1·2 |
| Occupation (%) | |
| Agriculture | 76·6 |
| Industry | 10·8 |
| Service | 5·0 |
| Cadre | 4·4 |
| Other | 5·2 |
| Geographic region of residence (%) | |
| Beijing/Shanghai/Tianjin | 9·1 |
| North-east | 10·3 |
| North | 11·9 |
| East | 23·9 |
| South | 21·3 |
| South-west | 9·2 |
| North-west | 14·2 |
| Ethnicity (%) | |
| Han | 89·2 |
| Minority | 10·8 |
| Total number of women | 27,655 |

to indicate if transformations are necessary and, if so, to suggest empirical parametric or non-parametric transformations for use in linear regression. In this application, the non-parametric transformations estimated by ACE are not used in their raw form, but functional approximations to them are found that are then used in the estimation process.

There are a number of other non-parametric methods for finding a variable transformation such as the generalized additive model (GAM) (Hastie & Tibshirani, 1990). ACE was chosen for two reasons. First, ACE can be used to transform both independent and dependent variables in a linear regression model, whereas other methods like GAM can only be applied to transform the independent variables. For example, it may be that a logarithmic transformation of the dependent variable is the appropriate one in a regression model, which can only be detected by ACE algorithm. Second, ACE is a fully automatic smoothing tool that generates diagnostic outputs and plots, making it easy to use and interpret. In addition, other non-parametric regression methods may require the pre-specification of the smoothing parameters, which is often difficult to determine. For example, to use GAM for variable transformations, smoothing parameters, or target-equivalent degrees of freedom, need to be specified in all the smoothing terms in the GAM model so that the local scoring algorithm can be performed, which may be a problem for an inexperienced data analyst.

The ACE algorithm is implemented using the *ace* and *avas* functions in the S-PLUS statistical package (Venables & Ripley, 1994). To apply the ACE algorithm to possibly recode an independent variable, an examination of the scatterplot of the original values against the ACE-transformed values indicates the nature of an approximately optimal transformation. For example, if the plot is judged linear, a linear model with the original data is appropriate; if not, it suggests what kind of transformation should be employed. Possibilities include power and log transformations, threshold and piecewise linear functions, categorical (i.e. piecewise constant) transformations, and quadratic or higher order polynomial terms. There are often a number of potential candidates for transformation of a variable. To select the best transformation, the BIC (Bayesian Information Criterion) is used because most models considered here are non-nested and this statistic has been demonstrated to be a good index for comparing non-nested models (Raftery, 1995). The BIC statistic, originally derived from a Bayesian approach to statistical inference, can be expressed as:

$$\mathrm{BIC} = N \log(1 - R^2) + p \log(N)$$

where $R^2$ is the adjusted square of the multiple correlation coefficient, $p$ is the number of independent variables in the model of interest and $N$ is the sample size, which in this case is 27,655. The smaller BIC is, the better the model.

To illustrate the use of the ACE algorithm, it was first applied to recode the year of birth of a woman (i.e. cohort). The next step shows how ACE is used to guide the transformation of this cohort variable, and how BIC is used to select its best transformation. The same procedure is used to recode other variables. The transformation of cohort is undertaken in two steps. First the ACE algorithm is used to get a transformed function of cohort, and this is plotted against the original year of birth (measured as years since 1900). The result is shown in Fig. 1. Secondly, based on this estimated pattern, alternative transformations of cohort are then tried and the best one according to BIC selected.

The ACE result in Fig. 1 suggests at least five possible parametric transformations for cohort: (1) a piecewise linear spline with a node at around 1960; (2) a simple straight line over the whole period; (3) a quadratic curve; (4) an exponential transformation; and (5) a logarithm transformation. The BIC statistics corresponding to these five transformation types are presented in Table 2.

Table 2 shows that all the proposed specifications are statistically significant according to their $F$-statistics. However, the relative goodness-of-fit of those transformations cannot be judged using these $F$-statistics because the models are non-nested. For this purpose, BIC statistics are used, which show that the squared cohort transformation fits best, although the degree of curvature is slight.

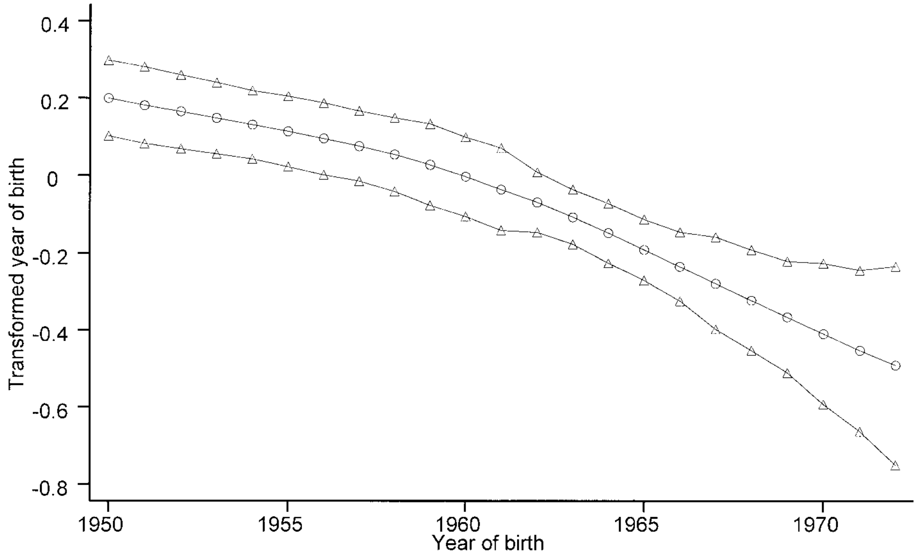


**Fig. 1.** ACE transformation of year of birth and 95% confidence intervals.

**Table 2.** Selecting the best transformation for year of birth

| Specifications | $p$ | $F$ | $R^2$ | BIC |
|---|---|---|---|---|
| Parametric specification | | | | |
| Linear | 1 | 397·0 | 0·0141 | −382·5 |
| Squared | 1 | 400·6 | 0·0142 | −385·3 |
| Linear+squared | 2 | 201·7 | 0·0143 | −377·9 |
| Exponential | 1 | 53·3 | 0·0019 | −42·4 |
| Logarithm | 1 | 392·3 | 0·0141 | −382·5 |
| Piecewise linear splines | | | | |
| Linear spline with node at 1959 | 2 | 201·8 | 0·0143 | −377·9 |
| Linear spline with node at 1960 | 2 | 201·6 | 0·0143 | −377·9 |
| Linear spline with node at 1961 | 2 | 201·1 | 0·0143 | −377·9 |

Sample size ($N$)=27,655.
$p$, degrees of freedom.
$F$, statistic for testing if regression model holds.
$R^2$, adjusted square of the multiple coefficient in the linear regression.
$\mathrm{BIC}=N\log(1-R^2)+p\log(N)$

As mentioned before, the ACE approach can also be applied to establish if transformation is necessary for the dependent variable. The ACE plot shows a linear relationship between untransformed and transformed age at menarche (not shown here), so no variable transformation is needed.

The coding of other variables using the ACE algorithm is now described. So far, the ACE-guided transformation has been considered just for one independent variable, namely year of birth of the woman, but the approach for more than one predictor variable is a straightforward extension of the bivariate case, allowing a matrix of independent variables to be transformed simultaneously. The following analysis is based on transformation of all the independent variables shown in Table 2. The plots of the transformed variables against the original ones are shown in Fig. 2.

Using analyses similar to those used for coding the period variable, the following choices were made: ‘educational attainment’ is measured as the reported highest level of educational achievement (Table 1). The plot of ACE transformation of this variable suggests non-linearity between educational achievement and menarcheal age. BIC statistics show that three categories (1 for illiterate/semiliterate; 2 for primary school and junior or senior middle school; and 3 for college or above) can best describe the effect of this variable on age at menarche.

‘Occupation’ refers to current or recent employment of the woman and was coded in five categories (Table 1). After a number of experiments, the lowest BIC score was obtained by simply reclassifying the occupation into two broad categories: industry/service/professional and agriculture/other.

Three municipal cities and 26 provinces were sampled in this survey. In order to simplify presentation of the region of residence variable, the three metropolitan cities were put into one group, and the 26 provinces into six large geographic areas as follows: 1, Beijing, Shanghai and Tianjin; 2, North; 3, North-east; 4, East; 5, South; 6, South-west; and 7, North-west. The ACE plot for the transformation of this variable suggests some possible regrouping: regions 2, 5 and 7 have about the same transformed values, and so do regions 3 and 4. BIC statistics show that four new categories form a satisfactory coding for representing regional differentials in age at menarche with this sample size: separate ones for regions 1 and 6; and combining regions 2, 5 and 7; and regions 3 and 4. While alternative groupings of regions might produce better statistical fits, this was outweighed by the benefits of having easy to interpret, well-defined geographical areas.

## Results

Table 3 gives descriptive statistics on the age at menarche of the sample of 27,655 ever-married women born in four successive cohorts. A decrease in mean menarcheal age is clearly recorded in these four cohorts. Standard deviation values show little change in the variability of the phenomenon through the time span examined, although these are rather higher than those of Lin *et al*. (1992). To understand the effects of socioeconomic variables on the decline in the recalled age at menarche, a multiple linear regression model analysis is now considered.

Table 4 presents a series of six linear regression models of age at menarche with increasing numbers of covariates that have been identified as likely to be associated with age at menarche. Model 1 is presented for completeness; Model 2 includes urban–rural residence; Model 3 presents estimates of educational level; Model 4 contains occupation; Model 5 includes estimates for all the socioeconomic variables;

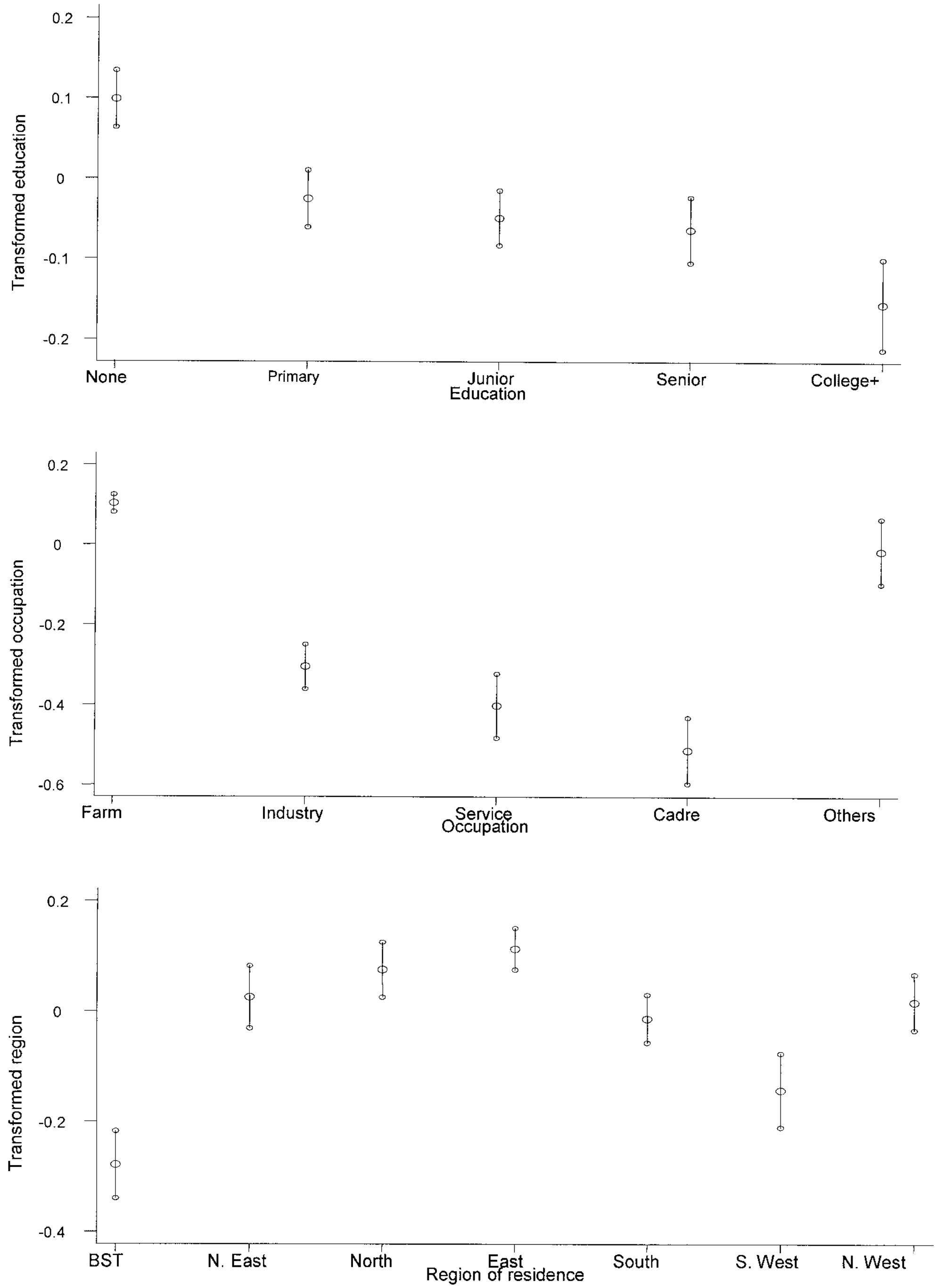


**Fig. 2.** ACE transformation plots for education, occupation and residence. Bar indicates 95% confidence interval of ACE estimates. Region of residence codes are as follows: BST, Beijing, Shanghai and Tianjin; N. East, North-east; S. West, South-west; N. West, North-west.

**Table 3.** Mean age of menarche by year of birth

| Year of birth | No. of women | Mean | SD |
|---|---|---|---|
| 1950–1954 | 7418 | 14·50 | 1·87 |
| 1955–1959 | 7702 | 14·39 | 1·89 |
| 1960–1964 | 8199 | 14·12 | 1·73 |
| 1965+ | 4336 | 13·95 | 1·58 |

and finally Model 6 presents the full model, including region of residence and ethnicity. These models are organized so as to address a number of important questions by comparison of different models.

First, a comparison of Model 1 with other models shows the extent to which the decline in age at menarche for women born in different cohorts can be explained by the changes in socioeconomic and other variables. The quadratic effect for the cohort variable representing the year of woman's birth in Table 4 suggests that there is no slowdown in the decline in menarcheal age for recent cohorts, and that the estimated decline for cohorts born between 1950 and 1972 is 1·21 years. This decline in national levels is very similar in magnitude to the 1·5 year decline in median menarcheal age for urban Beijing between 1962 and 1985 reported by Lin *et al.* (1992), based on two current status surveys and therefore referring to cohorts born around 1948 and 1972 respectively. Graham *et al.* (1999) found that the mean age at menarche decreased by 2·8 years over an approximate 40-year interval from about 1940 based on the data from two counties in Anhui Province. After adding the full set of independent variables to Model 1, the coefficient of the cohort variable in Model 6, as in all the intermediate models, remains almost unchanged and statistically significant, suggesting that the decline in the age at menarche by birth cohort cannot be explained by changes in these socioeconomic variables included in this study. Thus there are other factors which played an important role in lowering the age at menarche. The general improvements in standards of living in China over the last several decades, and especially better nutrition, are likely to have contributed to the menarcheal age decline.

Second, comparison of Models 2, 3 and 4 with Models 5 and 6 provides the means of assessing the extent to which the impact of each of the socioeconomic variables is altered by the inclusion of a fuller set of socioeconomic determinants of age at menarche. The results show that the magnitudes of the urban–rural residence, education and occupation coefficients are considerably reduced when they are introduced in the models simultaneously as these three variables are highly correlated. For example, the effect of urban–rural residence is more than halved, from 0·95 to 0·38 years, when education and occupation are included in Model 5.

Third, the final Model 6 shows the parameter estimates of the full set of socioeconomic covariates. Comparison of Models 2–6 shows that the socioeconomic variables selected in this study have a statistically significant and direct impact on menarcheal age after controlling for other variables. Even after controlling for the

**Table 4.** Estimated coefficients (and standard errors) for linear regression models of age at menarche[a]

| Variable | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 |
|---|---|---|---|---|---|---|
| Intercept | 16·98 (0·061) | 17·08 (0·062) | 17·46 (0·061) | 17·17 (0·072) | 17·08 (0·079) | 16·67 (0·086) |
| Squared year of birth (from 1900) | −0·00045 (0·00002) | −0·00042 (0·00002) | −0·00041 (0·00002) | −0·00043 (0·00002) | −0·00042 (0·00002) | −0·00042 (0·00002) |
| Rural residence (urban)[b] | | 0·95 (0·025) | | | 0·38 (0·043) | 0·36 (0·042) |
| Education (Illiterate/semiliterate)[b] | | | | | | |
| Primary/junior/senior | | | −0·19 (0·022) | | −0·17 (0·021) | −0·17 (0·021) |
| College and above | | | −1·55 (0·099) | | −0·74 (0·099) | −0·68 (0·098) |
| Occupation in industry/service/cadre (agriculture/other)[b] | | | | −1·06 (0·026) | −0·69 (0·045) | −0·63 (0·045) |
| Geographic region of residence (Beijing/Shanghai/Tianjin)[b] | | | | | | |
| North-east/South/North-west | | | | | | 0·49 (0·039) |
| North/East | | | | | | 0·65 (0·040) |
| South-west | | | | | | 0·26 (0·050) |
| Han ethnicity (minority)[b] | | | | | | −0·046 (0·015) |
| Model fit statistics[c] | | | | | | |
| $p$ | 1 | 2 | 3 | 2 | 5 | 9 |
| $R^2$ | 0·0142 | 0·0636 | 0·0244 | 0·0688 | 0·0754 | 0·0862 |
| $F$ | 400·9 | 940·0 | 231·0 | 1022·0 | 452·0 | 290·8 |
| BIC | −385·3 | −1796·8 | −641·1 | −1590·8 | −2113·9 | −2400·9 |

[a]All the coefficients are statistically significant to at least the 1% level.
[b]Omitted category is enclosed in parentheses.
[c]$p$, degrees of freedom; $R^2$, adjusted square of the multiple coefficient in the linear regression; $F$, statistic for testing if regression model holds; $\mathrm{BIC} = N \log(1 - R^2) + p \log(N)$; sample size $(N) = 27{,}655$.

generally superior educational and occupational composition of urban populations, living in an urban setting in developing countries such as China is likely to have a positive impact on lowering the age at menarche through a variety of factors, including higher consumption of animal protein and fat (Qu *et al.*, 2000) and more advanced medical facilities and public health systems in cities (Knight & Song, 1999), so it is no surprise that women in rural areas have a significantly higher age at menarche than their urban counterparts with otherwise similar characteristics.

The coefficients in Table 4 show a relationship between lower age at menarche and (subsequent) higher educational level of the woman, even when the effects of other variables, including occupation, are held constant. The result is consistent with the findings from a case study of two rural counties in Anhui Province (Graham *et al.*, 1999). Girls without any schooling or with little schooling had their menarche 0·68 years later than those with college education or above. This probably reflects the generally higher socioeconomic status of highly educated married women, since selection for young age at menarche among this group is unlikely.

Social class has always been an important variable in the studies of age at menarche (Mascie-Taylor & Boldsen, 1986). Women's subsequent occupation is also found to be an important correlate of their age at menarche. Women working in industry, service and professional areas had a significantly lower age at menarche than women working in other sectors, in this case, predominantly in agriculture. In terms of ethnicity, girls of national minorities have higher menarcheal age than those of the Han majority, who account for over 93% of the total population but a lower proportion of births, consistent with previous studies (Lin *et al.*, 1992; Quan, 1984; Lin & Yuan, 1989). Although individual level socioeconomic variables have been controlled for, this is probably because these national minorities tend to live in remote, underdeveloped frontier areas, where medical and other facilities are inferior.

There are also regional differentials in age at menarche. Compared with the three municipal cities (Beijing, Tianjin and Shanghai), the other regions have a higher age at menarche, particularly the three regions North-east, South and North-west, which have the highest menarcheal age, 0·66 years higher than the value of the three municipal cities. To examine whether the rate of decline was greater in different areas, an interaction term between cohort and place of residence was fitted, but there was no statistically significant evidence that the declines were different in different parts of the country. Although noted that there is a genetic correlation of 0·2–0·3 between ages of menarche of mothers and daughters, girls living in different Chinese regions represent a situation where a relatively homogeneous population lives in different sociogeographic environments. Therefore, these results suggest (but do not prove) that the regional differences of age at menarche are likely to be mainly due to environmental influences.

## Conclusion

This study analyses some socioeconomic determinants of age at menarche in China using the data from the Two-per-Thousand Chinese Fertility Survey. A linear regression model was employed to examine the trends and differentials in age at menarche for women born between 1950 and 1973. To better describe the relationship between menarcheal age and its determinants, the ACE algorithm was used to efficiently transform the independent variables. This study has demonstrated the procedures and usefulness of the ACE algorithm for helping to understand the relationship between the independent variables and age at menarche. The successful modelling of a data set and appropriate coding of a variable are part science, part statistical methods, and part experience and common sense. The advantage of applying the ACE algorithm is that it provides a number of plausible alternative candidates, from which an appropriate specification can be chosen, either based on a statistical criterion (such as BIC or chi-square statistics) or based on past experience or scientific knowledge. It is hoped that this analytical tool will provide new encouragement for the empirical modelling of biological phenomena.

Chinese girls born in the third quarter of the twentieth century have experienced a steady decline in age at menarche. This study has shown how far this declining trend is attributable to changes in socioeconomic conditions, and found that the decline in the age at menarche for women born between 1950 and 1973 cannot be explained by changes in the variables included in the study. Thus, social changes that have occurred in China, but are not measured in surveys such as this one, have played an important role in reducing average menarcheal age. Those social changes are likely to be additional influences accompanying socioeconomic development.

This study shows that after appropriate transformations, all the socioeconomic variables selected have statistically significant and direct influence on menarcheal age. Women living in rural areas, with less education, engaging in agricultural work, and being members of national minorities, had a higher average age at menarche. Current trends in the socioeconomic conditions of the Chinese population suggest that the average age at menarche will continue to fall.

## References

Breiman, L. & Friedman, J. H. (1985) Estimating optimal transformations for multiple regression and correlation. *J. Am. statist. Ass.* **80**, 580–597.

Coale, A. J. & Banister, J. (1994) Five decades of missing females in China. *Demography* **31**(3), 459–479.

De Veaux, R. D. (1989) Finding transformation for regression using the ACE algorithm. *Sociol. Meth. Res.* **18**, 327–359.

Eveleth, P. B. (1986) Timing of menarche: secular trend and population differences. In: *School-Age Pregancy and Parenthood, Biological Dimensions*, pp. 39–52. Edited by J. B. Lancaster & B. A. Hamburg. Aldine, New York.

Frisch, R. E. (1985) Fatness, menarche, and female fertility. *Perspect. Biol. Med.* **28**, 611–633.

Garn, S. M. (1980) Continuities and change in maturational timing. In: *Constancy and Change in Human Development*, pp. 113–162. Edited by O. G. J. Brim & J. Kagan. Harvard University Press, Cambridge, MA.

Graham, M. J., Larsen, U. & Xu, X. (1999) Secular trend in age at menarche in China: a case study of two rural counties in Anhui Province. *J. biosoc. Sci.* **31**, 257–267.

Hastie, T. J. & Tibshirani, R. J. (1990) *Generalized Additive Models*. Chapman and Hall, New York.

Hediger, M. L. & Stine, R. A. (1987) Age at menarche based on recall Information. *Ann. hum. Biol.* **14**(2), 133–142.

Knight, J. B. & Song, L. (1999) *The Rural–Urban Divide: Economic Disparities and Interactions in China*. Oxford University Press, Oxford.

LAVELY, W. R. (1991) China unveils its monumental two per thousand fertility survey. *Asian and Pacific Popul. Forum* **5**(4), 89–92.

LI, H. (1991) Introduction to the National Survey of Fertility and Birth Control. In: *Proceedings of the National Conference on China's 1988 Two-per-Thousand Fertility and Birth Control Survey*, pp. 2–9. Hangzhou, China.

LIN, W. S., CHEN, A. C. N., SU, J. Z. X., ZHU, F. C., XING, W. H., LIN, J. Y. & YE, G. S. (1992) The menarcheal age of Chinese girls. *Ann. hum. Biol.* **19**(5), 503–512.

LIN, W. S. & YUAN, J. (1989) Menarche in Chinese girls and its relationships with physical development. *Sport Sci.* **9**(3), 13–17 (in Chinese).

MASCIE-TAYLOR, C. G. N. & BOLDSEN, J. L. (1986) Recalled age of menarche in Britain. *Ann. hum. Biol.* **13**(3), 253–257.

QU, J. B., ZHANG, Z. W., SHIMBO, S., LIU, Z. M., CAI, X. C., WANG, L. Q., WATANABE, T., NAKATSUKA, H., MATSUDA-INOGUCHI, N., HIGASHIKAWA, K. & IKEDA, M. (2000) Nutrient intake of adult women in Jilin Province, China, with special reference to urban–rural differences in nutrition in the Chinese continent. *Eur. J. clin. Nutr.* **54**(10), 741–748.

QUAN, Q. L. (1984) The study on physical pattern in menstruating and non-menstruating girls. C*hinese School Hlth* **5**(1), 36–39 (in Chinese).

RAFTERY, A. E. (1995) Bayesian model selection in social research (with Discussion by A. Gelman, D. B. Rubin & R. M. Hauser, and a Rejoinder). *Sociol. Methodol.* **25**, 111–195.

SANDLER, D. P., WILCOX, T. O. & HONEY, L. F. (1984) Age at menarche and subsequent reproductive events. *Am. J. Epidemiol.* **119**, 765–774.

TANNER, J. M. & EVELETH, P. B. (1975) Variablity between populations in growth and development at puberty. In: *Puberty: Biological and Psychosocial Components*, pp. 256–273. Edited by S. R. Berenberg. Stenfert Kroese Publishers, Leiden.

WANG, D. & MURPHY, M. (1998) Covariates of infant mortality in China: an exploratory approach. *Social Biol.* **45**(1–2), 21–38.

WYSHAK, G. (1983) Age at menarche and unsuccessful pregnancy outcome. *Ann. hum. Biol.* **10**(1), 69–73.

WYSHAK, G. & FRISCH, R. E (1982) Evidence for a secular trend in age of menarche. *New Eng. J. Med.* **306**(17), 1033–1035.

URDY, J. R. & CLIQUET, R. L. (1982) Cross-cultural examination of the relationship between ages and menarche, marriage, and the first birth. *Demography* **19**(1), 53–63.

VENABLES, W. N. & RIPLEY, B. D. (1994) *Modern Applied Statistics with S-Plus.* Springer-Verlag Inc., New York.